



**POLITECHNIKA WARSZAWSKA**

Wydział Elektroniki i Technik Informacyjnych

Instytut Informatyki

**Michał Haponiuk**

nr albumu: 249371

PRACA DYPLOMOWA INŻYNIERSKA

**Przeglądarka genomu**

Praca wykonana pod kierunkiem

dr hab. inż. Robert Nowak

Warszawa, 15 października 2015

# Spis treści

<b>1. Przeglądarki genomów</b>	<b>2</b>
1.1 Mapowanie genomów	3
1.1.1 Mapy genetyczne	4
1.1.2 Mapy fizyczne	4
1.2 Historia	4
1.3 Oprogramowanie i narzędzia	5
1.3.1 Bazy danych	6
1.3.2 Usługi internetowe	6
1.4 Najpopularniejsze przeglądarki	7
1.4.1 Dostęp	7
1.4.2 Architektura	7
1.4.3 Przykłady	7
<b>2. Projekt i implementacja</b>	<b>9</b>
2.1 Architektura	10
<b>Bibliografia</b>	<b>13</b>

# 1. Przeglądarki genomów

Bioinformatyka jako dziedzina nauki łączy w sobie informatykę i matematykę w celu przetwarzania danych biologicznych. Szybki rozwój genomiki i biologii molekularnej doprowadził do powstania ogromnej ilości informacji. Eksperymenty biologiczne zaczęły generować ogromne ilości danych, które okazały się trudne do przechowywania i „ręcznego” przetwarzania. Biolodzy w swojej pracy byli zmuszeni do skorzystania z komputerów i informatyki na etapach gromadzenia, wydobywania oraz przetwarzania informacji zawartych w DNA<sup>1</sup> (rys.1.1), RNA<sup>2</sup> i białkach.



Rysunek 1.1. Prosty model DNA.

Matematyka wraz z informatyką dostarczają niezbędnych narzędzi do rozwiązywania problemów związanych z biologią. Pociąga to za sobą powstanie i zarządzanie zaawansowanymi bazami danych, w których mogą być przechowywane i efektywnie eksplorowane dane biologiczne. Kolejną ważną częścią bioinformatyki jest aspekt obliczeniowy, czyli wykorzystanie komputerów do przetwarzania danych. Matematyka pozwala przy pomocy równań zamodelować pewne zjawiska, ale przede wszystkim do skonstruowania rozwiązań potrzebne są efektywne algorytmy i programy komputerowe.

Przeglądarki genomów są narzędziem, którym posługuje się bioinformatyka sekwencji. Dzięki nim możliwe jest katalogowanie informacji biologicznych, przeszukiwanie baz danych sekwencji, porównywanie genomów i sekwencjonowanie. Otwarte podejście do stosowania narzędzi bioinformatycznych znacząco skraca czas planowania i przeprowadzania eksperymentów oraz oszczędza pieniądze.

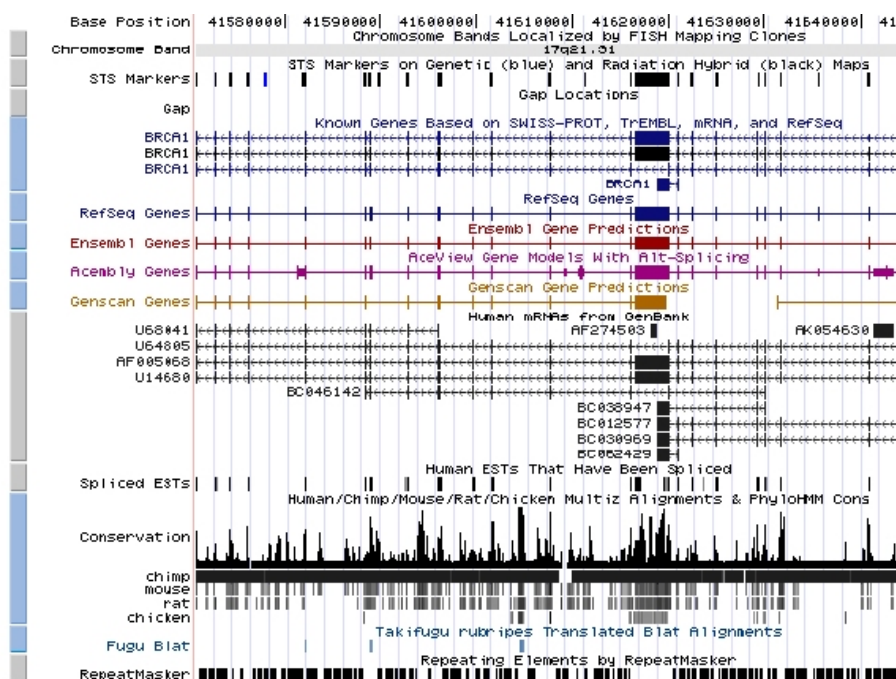
Obecne przeglądarki są w formie graficznych interfejsów, wyświetlających dane genomów z biologicznych baz danych wraz z adnotacjami. Umożliwiają naukowcom wizualizacje

---

<sup>1</sup> Kwas deoksyrybonukleinowy

<sup>2</sup> Kwas rybonukleinowy

całych genomów z dodatkowymi informacjami pozyskiwanymi zwykle z wielu źródeł<sup>3</sup>. Różnią się one od zwykłych biologicznych baz danych, gdyż dane prezentowane są w graficznym formacie ze współrzędnymi genomów naniesionymi na oś. Rysunek 1.2 ilustruje typową wizualizację genomu.



Rysunek 1.2. Fragment typowej wizualizacji genomu.

## 1.1 Mapowanie genomów

Nieodzownym elementem tworzenia przeglądarki genomów jest sporządzenie map przedstawiających położenie poszczególnych genów na chromosomie. Wyróżnia się *mapy genetyczne*, określające odległości pomiędzy charakterystycznymi sekwencjami - markerami, oraz *mapy fizyczne*, przedstawiające rzeczywistą pozycję poszczególnych genów. Informacje zawarte w mapach fizycznych i genetycznych uzupełniają się nawzajem i stanowią podstawę dalszych badań nad organizacją i budową chromosomów.

<sup>3</sup> Dane porównawcze (drzewa filogenetyczne, zmienności), laboratoryjne badania analityczne (sekwencje, skany), odnośniki do innych baz, przetworzone dane laboratoryjne.

### 1.1.1 Mapy genetyczne

Mapy genetyczne na podstawie analizy sprzężeń genetycznych pozwalają ustalić rozmieszczenie genów i innych charakterystycznych sekwencji (markerów) w genomie oraz określają odległości genetyczne pomiędzy nimi. Jednostką w tych mapach jest centymorgan<sup>4</sup> (cM).

### 1.1.2 Mapy fizyczne

Mapy fizyczne wykorzystują techniki biologii molekularnej do bezpośredniej lokalizacji różnych typów sekwencji DNA w genomie. Stosowaną jednostką są tu pary zasad<sup>5</sup> (pz). Tablica 1.1 przedstawia dwa komplementarne łańcuchy, oba mające długość 15 pz.

Łańcuch 1	A	T	C	G	A	T	T	G	A	G	C	T	C	T	A
Łańcuch 2	T	A	G	C	T	A	A	C	T	C	G	A	G	A	T

Tablica 1.1. Komplementarność zasad w przykładowym łańcuchu DNA.

## 1.2 Historia

Początki bioinformatyki sięgają lat osiemdziesiątych, kiedy to utworzono w Stanach Zjednoczonych bazę danych zwaną *GenBankiem*. Wraz z coraz częstszymi badaniami różnych organizmów i sekwencjonowaniem ich przez badaczy, Amerykański Departament Energii zdecydował się na opracowanie ww. programu, aby zbierać i sekwencjonować krótkie odcinki DNA.

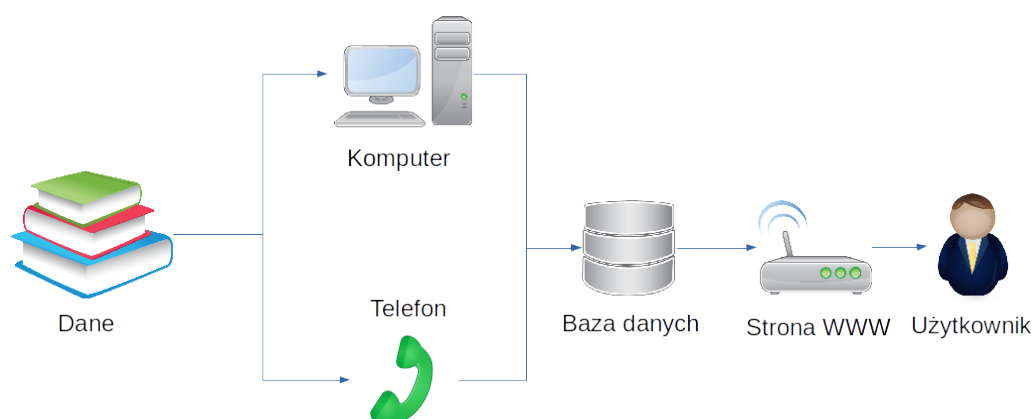
Początkowo użytkownicy wprowadzali sekwencje DNA do systemu używając specjalnych klawiatur posiadających jedynie cztery klawisze - A, C, T oraz G. Z biegiem czasu opracowano specjalne protokoły przekazywania danych przyspieszające cały proces. Naukowcy mogli zadzwonić do *GenBanku* i używając komputera byli w stanie bezpośrednio wprowadzić dane o kolejnych fragmentach genomu do bazy (rys.1.3). Rozwój technologii internetowych diametralnie zmienił sposób przekazywania informacji pomiędzy ośrodkami naukowymi. Po utworzeniu serwisu WWW, uczeni z całego świata otrzymali bezpłatny dostęp do zgromadzonych danych.

<sup>4</sup> Jeden cM to taka odległość pomiędzy dwoma loci, że szansa na ich rozdzielenie w procesie rekombinacji genetycznej w ciągu jednego pokolenia (podczas jednorazowego wydarzenia, np. crossing-over) wynosi 1%. Centymorgany nie odzwierciedlają bezwzględnej odległości pomiędzy obszarami chromosomu zajmowanymi przez gen, ponieważ częstość rekombinacji jest różna w różnych rejonach chromosomów, na różnych chromosomach oraz w różnych organizmach.

<sup>5</sup> Para zasad to dwie komplementarne zasady nukleotydów dwóch różnych nici kwasu nukleinowego, połączone wiązaniem wodorowym.

Zarząd *GenBanku* przeniesiono do NCBI<sup>6</sup> w National Institutes of Health. Gdy w 1990 roku oficjalnie wystartował Projekt Poznania Ludzkiego Genomu, ilość informacji napływająca do *GenBanku* rosła w tempie wykładniczym. Wraz z opracowaniem wydajnych metod sekwencjonowania wykorzystujących komputery, automatyczne sekwencjonery i zagadnienia robotyki, proces gromadzenia danych niesamowicie przyspieszył.

Równolegle prywatne firmy również zajęły się podobnymi badaniami tworząc własne ogromne bazy danych. Obecnie największe korporacje są w stanie odczytać w ciągu jednego dnia zapis genetyczny liczący około 20 milionów par zasad DNA. Ilość danych jaką uzyskano z pełnej sekwencji genomu ludzkiego to ponad 50 TB<sup>7</sup>



Rysunek 1.3. Historyczny schemat przepływu informacji o sekwencjach.

### 1.3 Oprogramowanie i narzędzia

Spektrum oprogramowania dostępnego dla bioinformatyki jest bardzo szerokie. Rozpoczyna się od prostych narzędzi z interfejsem terminalowym, a kończy na usługach internetowych i złożonych programach z możliwościami rozbudowanej interakcji z użytkownikiem.

Znaczna część narzędzi jest dostępna w formie darmowej lub otwartego oprogramowania. Następstwem zapotrzebowania na nowe algorytmy służącego do analizy był dynamiczny rozwój darmowych programów. Dzięki ideologii otwartego kodu źródłowego, grupy badawcze z całego świata mogą przyczynić się do tworzenia coraz to lepszych rozwiązań. Istotną zaletą darmowego oprogramowania jest tworzenie pewnych standardów i integrowanie bioinformatycznych danych.

<sup>6</sup> National Center for Biotechnology Information

<sup>7</sup> Taka ilość informacji zajęłaby 80 tys. płyt kompaktowych.

### 1.3.1 Bazy danych

Bazy danych używane w przeglądarkach genomów zwykle są ogromnymi zbiorami informacji, w większości składającymi się z danych sekwencji z grupy kwasów nukleinowych. Z powodu, że są konieczne przy badaniach naukowych tworzenie ich jest jednym z najistotniejszych zagadnień w bioinformatyce. Dane przechowywane w biologicznych bazach nie są niezmiennie i ostateczne. Wszystkie sekwencje są wynikiem badań o jakiejś dokładności. Okazuje się, że czasami dane zgromadzone w bazach zawierają jednak nieprawidłowości.

Możliwość szybkiego i precyzyjnego wyszukiwania pożądaných informacji jest jedną z ważniejszych cech biologicznych baz danych. Najczęściej użytkownik oprócz obróbki danych ma również możliwość połączenia się przez internet z innymi bazami. Charakteryzują się zazwyczaj, także małą redundancją danych. By ułatwić pracę, uczeni często mają możliwość wyboru sposobu wyświetlania wyszukanych danych oraz wydruku bądź zapisu ich.

Pierwsze bazy danych zazwyczaj budowali biolodzy molekularni i biochemicy. W dzisiejszych czasach zwraca się uwagę, aby były one czytelne nie tylko dla specjalistów, ale i dla badaczy z nauk pokrewnych biologii.

### 1.3.2 Usługi internetowe

Dla wielu zastosowań bioinformatycznych zostały opracowane interfejsy wykorzystujące protokoły SOAP<sup>8</sup> i REST<sup>9</sup>. Dzięki takim rozwiązaniom, aplikacja działająca w jednej części świata jest w stanie używać algorytmów, danych i zasobów obliczeniowych serwerów podpiętych do sieci w innych częściach świata. Największe udogodnienia wynikają z faktu, że użytkownicy końcowi nie muszą zajmować się utrzymywaniem baz danych oraz oprogramowaniem. Usługi bioinformatyczne zostały podzielone na trzy podstawowe filary: dotyczące poszukiwania sekwencji, wielokrotnego przyrównywania sekwencji oraz analizy biologicznych sekwencji. Podział nastąpił z inicjatywy Europejskiego Instytutu Informatyki (EBI).

---

<sup>8</sup> Simple Object Access Protocol - protokół komunikacyjny, korzystający z XML do kodowania wywołań.

<sup>9</sup> Representational State Transfer - styl architektury oprogramowania wywiedziony z doświadczeń przy pisaniu specyfikacji protokołu HTTP dla systemów rozproszonych, wykorzystuje m.in. jednorodny interfejs oraz bezstanową komunikację.

## 1.4 Najpopularniejsze przeglądarki

### 1.4.1 Dostęp

Na rynku obecnie dostępne jest wiele przeglądarek. Dzięki ogólnosiwiatowej sieci komputerowej są szeroko używane przez badaczy. Nie zawsze jednak możemy używać przeglądarek *online*. Możemy nie chcieć udostępniać danych na zewnątrz, albo możemy posiadać ich za dużo by przesłać je na serwer z innego krańca świata. Lekarstwem na ten problem są przeglądarki *offline* wyświetlające dane z dysku, ale pobierające kontekst opisu genomu ze zdalnego serwera DAS<sup>10</sup>.

### 1.4.2 Architektura

Różnią się niekiedy architekturą. Jedne są scentralizowane, inne stawiają na rozproszony system przechowywania baz danych. Oba podejścia charakteryzują się wadami i zaletami, w zależności od rozpatrywanego punktu widzenia. Projekty zaprzęgające dużą część społeczności naukowej wymagają wysiłku od dużej liczby osób, często powodując wygasanie entuzjazmu w pracy oraz wymaga praktycznie ciągłej aktualizacji. Specjalizowane bazy danych z silną kuratelą zmagają się niekiedy z problemami finansowymi. Sprawowanie ciągłej kontroli nad systemem również opóźnia dostęp do danych. Praktyka pokazuje, że tylko nieliczne bazy danych rozwijające się w zamkniętych społecznościach mają szansę dobrze funkcjonować, co rzutuje na popularność użytkowania przeglądarek genomów. Scentralizowane archiwa posiadają niekiedy kiepskie, niespójne albo wręcz żadne adnotacje<sup>11</sup> genomów.

### 1.4.3 Przykłady

- *UCSC browser* autorstwa Jima Kenta, napisana w większości w języku C w 2000 roku. Udostępniana darmowo dla akademickich zastosowań. Dla komercyjnych zastosowań wymagana licencja *Kent Informatics*.
- *Gbrowse* - największa część kodu to Perl, z czasem coraz więcej java-scriptu. Prace rozpoczęto w 2002 roku, funkcjonuje na licencji *PERL artistic license*.

---

<sup>10</sup> *Distributed Annotation System* to protokół wymiany danych stworzony w celu publikowania oraz integracji danych biologicznych, w systemie rozproszonym.

<sup>11</sup> Adnotacja jest daną powiązaną z jakimś elementem (np.sekwencją), dostarczającą nowych informacji na temat tego elementu.



- *ENSEMBL* - funkcjonuje na licencji Apache, dopuszczającej użycie kodu źródłowego zarówno na potrzeby wolnego oprogramowania, jak i zamkniętego oprogramowania komercyjnego. Znaczna część kodu napisana w Perlu, schemat bazy danych w MySQL.
- *Integrated Genome Browser* - projekt rozpoczęty przez firmę *Affymetrix*, jednak porzucony przez nią i rozwijany w środowisku akademickim. Przeglądarka napisana w Javie, obecnie na prawach Academic free license.

## 2. Projekt i implementacja

Aplikacje wykorzystywane do analizy danych genetycznych, korzystają zazwyczaj z ogromnych woluminów i złożonych algorytmów. Wysoka wydajność, elastyczność oraz interfejs użytkownika z wykorzystaniem przeglądarki internetowej to cechy aplikacji, które można osiągnąć łącząc ze sobą kilka języków programowania. [14]

Dane genetyczne przedstawiane są typowo w postaci zbiorów łańcuchów znaków, gdzie każdy ciąg jest sekwencją symboli z danego alfabetu. Reprezentacja łańcuchowa zwana pierwszorzędową (rys.2.1a), odzwierciedla fakt, że cząsteczki przechowujące informacje genetyczne (DNA i RNA) są biopolimerami<sup>1</sup> nukleotydów<sup>2</sup>, podczas gdy białka<sup>3</sup> są łańcuchami polipeptydowymi<sup>4</sup>. By zrozumieć wzajemne relacje pomiędzy nukleotydami a aminokwasami należy przyjrzeć się drugo, trzecio oraz czwartorzędowym strukturom. Struktura drugorzędowa (rys.2.1b) obejmuje wiązania wodorowe pomiędzy nukleotydami DNA i RNA oraz wiązania wodorowe pomiędzy peptydami w proteinach, określające przestrzenne ułożenie cząstek. Trzeciorzędowa struktura (rys.2.1c) określa położenie atomów w trójwymiarowej przestrzeni, natomiast czwartorzędowa (rys.2.1d) opisuje wyższy stopień organizacji wyjaśniający sposób łączenia się struktur trzeciorzędowych. [14, 15, 16, 17, 18, 19]

Wzbogacając model o informacje opisujące zależności pomiędzy ciągami a podciągami, możliwa jest analiza podobieństwa z wielu perspektyw. Ponadto, dane te uzupełnione o opisy czytelne dla człowieka ułatwiają zrozumienie znaczeń sekwencji oraz funkcji jakie pełnią. [14]

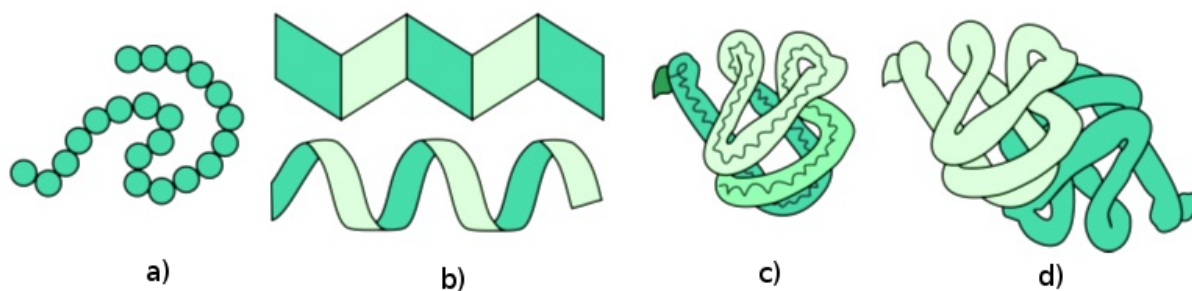
---

<sup>1</sup> Polimer występujący naturalnie w organizmach żywych, produkowany przez nie.

<sup>2</sup> Podstawowy składnik strukturalny kwasów nukleinowych.

<sup>3</sup> Białka, wielocząsteczkowe biopolimery.

<sup>4</sup> Łańcuchy cząstek aminokwasów połączonych wiązaniem peptydowym, po uzyskaniu odpowiedniej struktury przestrzennej może stać się funkcjonalnym białkiem.



Rysunek 2.1. Rzędowość struktury białek: pierwszorzędowa (a), drugorzędowa (b), trzeciorzędowa (c), czwartorzędowa (d).

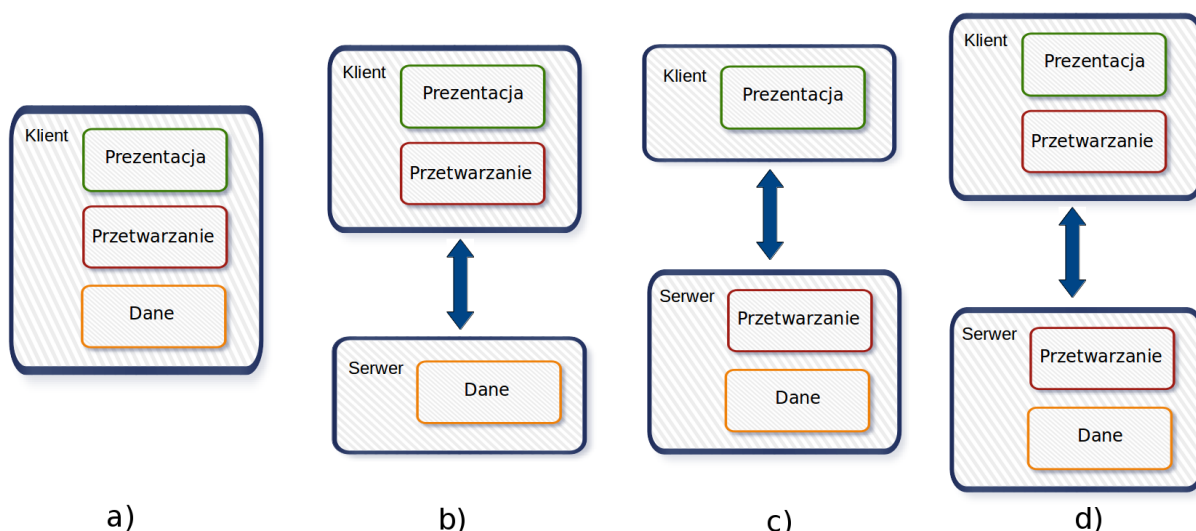
Charakterystyczną cechą programów komputerowych korzystających z danych genetycznych jest konieczność analizy ogromnych zbiorów danych za pomocą skomplikowanych algorytmów, co oznacza, że wysoka wydajność jest kluczowa. Różnorodność systemów i użytkowników sprawia, że przenośność oprogramowania jest równie ważna. Badacze preferują korzystanie z aktualizowanych automatycznie, graficznych interfejsów użytkownika dostępnych z poziomu przeglądarek internetowych. [14]

Inżynierowie starają się rozwiązywać występujące na globie praktyczne problemy. Naukowcy próbujący opisywać i wyjaśniać świat są coraz bardziej zaangażowani w rozwój oprogramowania. W celu przyspieszenia procesu tworzenia aplikacji oraz uniknięcia typowych błędów, badacze powinni korzystać ze zbioru metod, technik i technologii, którymi dysponuje dziedzina nauki nazywana inżynierią oprogramowania. Dzięki sprawdzonym praktykom możliwe staje się projektowanie systemów spełniających stawiane im wymagania. [14, 20]

## 2.1 Architektura

Aplikacja została zrealizowana w trójwarstwowej architekturze, gdzie warstwa prezentacji, przetwarzania oraz danych są wyraźnie odseparowane od siebie. Użycie wielowarstwowego modelu sprawia, że programy komputerowe są przenośne i wielorazowego użytku. Idea ta odzwierciedla fakt, że poszczególne moduły pełnią różne role w systemie. Warstwy pomagają oddzielić od siebie różne podsystemy, ułatwiając jednocześnie utrzymanie dobrze zorganizowanego, czytelnego kodu. [14]

Oprogramowanie zostało zbudowane w oparciu o swobodnie dostępny framework *bioweb* przystosowany do analizy danych genetycznych używający C++, Python (django), JavaScript (AngularJS) i kilku innych bibliotek. *Bioweb* został niejednokrotnie użyty do tworzenia aplikacji przetwarzających dane genetyczne skracając czas i koszty pracy do minimum. [14]



Rysunek 2.2. Modele aplikacji trójwarstwowej klient serwer: desktopowa (a), z serwerem bazodanowym (b), z cienkim klientem (c), aplikacja webowa (d).

Podczas projektowania aplikacji trójwarstwowej należy rozpatrzyć cztery możliwe modele: desktopowy (rys.2.2a), z serwerem bazodanowym (rys.2.2b), z cienkim klientem (rys.2.2c), oraz aplikację webową(rys.2.2d). Aplikacja desktopowa nie jest dobrym rozwiązaniem w projektowanym systemie, ponieważ nie zapewnia dostępu dla wielu użytkowników. Jedynie osoba korzystająca w danej chwili ze stacji roboczej mogłaby używać systemu. Z powodu braku scentralizowanego serwera bazodanowego ciężkim zadaniem stałoby się umożliwienie współpracy zespołom naukowym. Z racji łatwo dostępnej ogólnościwiatowej sieci komputerowej użytkowanie systemów w trybie offline staje się bardzo rzadkie. Bazy danych sekwencji są ogólnodostępne przez Internet, przez co połączenie do sieci okazuje się koniecznością przy badaniu danych genetycznych. [14]

Architektura aplikacji z udostępnioną współdzieloną bazą danych i modułem przetwarzania znajdującym się na maszynie klienta również nie jest najlepszą recepturą. Rozwiązanie to wymusza na kliencie posiadanie wysoko wydajnej maszyny w celu dokonywania skomplikowanych obliczeń na danych. Kolejnym problemem jest potrzeba aktualizacji oprogramowania po stronie klienta dla wielu platform co znacznie wydłuża czas i zwiększa koszty utrzymania systemu. [14]

Umiejscowienie modułów obliczeniowych na serwerze daje możliwość wykonywania obliczeń klientom osadzonym na różnych platformach, co zdecydowanie zmniejsza koszty produkcji i utrzymania oprogramowania. Istotną cechą serwera jest jego odpowiednio wysoka moc obliczeniowa. Od niej bowiem zależy, w jakim czasie uzyskamy wyniki pożądaných operacji. Takie rozwiązanie sprawia, że komputery klienckie nie muszą być wyposażone w wysoko

wydajny sprzęt elektroniczny. Odpowiednio wysokiej klasy sprzęt serwerowy gwarantuje stosunkowo szybkie wykonywanie obliczeń mimo zastosowania powolnego klienta. Optymalne rozmieszczenie modułów przedstawiają model aplikacji z cienkim klientem (rys.2.2c) oraz model aplikacji internetowej (rys.2.2d). [14]

Wariant aplikacji internetowej ma kilka zalet w porównaniu do systemu z cienkim klientem. Dobra taktyką jest by niektóre fragmenty serwerowej warstwy przetwarzania były po stronie klienta. Odciążenie serwera z operacji takich jak generowanie rysunków, grafów, walidacja danych wprowadzanych przez użytkownika usprawni komunikację między maszynami. Dzięki zastosowaniu technologii takich jak JavaScript/HTML5 wachlarz operacji wykonywanych przez graficzny interfejs użytkownika powiększa się. Prostszy również staje się aktualizacja oprogramowania klienta, ponieważ jego moduły są zawsze pobierane podczas inicjalizacji. [14]

# Bibliografia

- [1] Wikipedia (PL) *Bioinformatyka*  
<https://pl.wikipedia.org/wiki/Bioinformatyka>, 15 sierpnia 2015
- [2] Jan Paweł Jastrzębski *Wykład 2 bioinformatyka 2007/2008*  
[http://ebiolog.pl/graf/Wyklady/biotech/W2/Bioinformatyka\\_W2.pdf](http://ebiolog.pl/graf/Wyklady/biotech/W2/Bioinformatyka_W2.pdf),  
16 sierpnia 2015
- [3] Wikipedia (PL) *Centymorgan*  
<https://pl.wikipedia.org/wiki/Centymorgan>, 15 sierpnia 2015
- [4] e-Biotechnologia *Mapowanie genomów*  
<http://www.e-biotechnologia.pl/Artykuly/Mapowanie-genomow>,  
15 sierpnia 2015
- [5] e-Biotechnologia, Encyklopedia *Para zasad*  
<http://www.e-biotechnologia.pl/encyklopedia/articles.php?lng=pl&pg=175>, 15 sierpnia 2015
- [6] Wikipedia (EN) *Genome browser*  
[https://en.wikipedia.org/wiki/Genome\\_browser](https://en.wikipedia.org/wiki/Genome_browser), 16 sierpnia 2015
- [7] Uniwersytet Warmińsko Mazurski w Olsztynie *Historia Bioinformatyki...*  
<http://www.uwm.edu.pl/wnz/KBZ/Lipazy/bioinfo-historia.html>
- [8] Wikipedia (PL) *Kwas rybonukleinowy*  
[https://pl.wikipedia.org/wiki/Kwasy\\_rybonukleinowe](https://pl.wikipedia.org/wiki/Kwasy_rybonukleinowe) 16 sierpnia 2015
- [9] Wikipedia (PL) *SOAP*  
<https://pl.wikipedia.org/wiki/SOAP>, 17 sierpnia 2015
- [10] Wikipedia (PL) *REST*  
[https://pl.wikipedia.org/wiki/Representational\\_State\\_Transfer](https://pl.wikipedia.org/wiki/Representational_State_Transfer),  
17 sierpnia 2015
- [11] Anna Leśniewska *Przeglądarki genomowe (2)*  
[http://www.cs.put.poznan.pl/alesniewska/BABD/W\\_2015-05-06.pdf](http://www.cs.put.poznan.pl/alesniewska/BABD/W_2015-05-06.pdf),  
17 sierpnia 2015
- [12] Wikipedia (PL) *Apache License*  
[https://pl.wikipedia.org/wiki/Apache\\_License](https://pl.wikipedia.org/wiki/Apache_License), 17 sierpnia 2015

- [13] Bartek Wilczyński *Architektura dużych projektów bioinformatycznych*  
[http://www.cs.put.poznan.pl/alesniewska/BABD/W\\_2015-05-06.pdf](http://www.cs.put.poznan.pl/alesniewska/BABD/W_2015-05-06.pdf),  
17 sierpnia 2015
- [14] Robert M. Nowak *Polyglot programming the applications to analyze genetic data* BioMed  
Research International, vol. 2014, doi:10.1155/2014/253013, [http://www.hindawi.com/  
journals/bmri/2014/253013](http://www.hindawi.com/journals/bmri/2014/253013) 10 października 2015
- [15] Wikipedia (PL) *Biopolimery*  
<https://pl.wikipedia.org/wiki/Biopolimery>, 11 października 2015
- [16] Wikipedia (PL) *Nukleotydy*  
<https://pl.wikipedia.org/wiki/Nukleotydy>, 11 października 2015
- [17] Wikipedia (PL) *Białka*  
<https://pl.wikipedia.org/wiki/Bia%C5%82ka>, 11 października 2015
- [18] Wikipedia (PL) *Struktura drugorzędowa białka*  
[https://pl.wikipedia.org/wiki/Struktura\\_drugorz%C4%99dowa\\_bia%C5%  
82ka](https://pl.wikipedia.org/wiki/Struktura_drugorz%C4%99dowa_bia%C5%82ka), 11 października 2015
- [19] e-Biotechnologia *Budowa białek*  
<http://www.e-biotechnologia.pl/Artykuly/budowa-bialek>,  
11 października 2015
- [20] Krzysztof Sacha *Inżynieria oprogramowania* PWN, Warszawa 2010